D136–D140 *Nucleic Acids Research, 2011, Vol. 39, Database issue* *Published online 23 October 2010*
doi:10.1093/nar/gkq964

YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface

Dário Abdulrehman[1,2], Pedro Tiago Monteiro[1], Miguel Cacho Teixeira[2,3], Nuno Pereira Mira[2,3], Artur Bastos Lourenço[2,3], Sandra Costa dos Santos[2,3], Tânia Rodrigues Cabrito[2,3], Alexandre Paulo Francisco[1,2], Sara Cordeiro Madeira[1,2], Ricardo Santos Aires[1,2], Arlindo Limede Oliveira[1,2], Isabel Sá-Correia[2,3] and Ana Teresa Freitas[1,2,*]

[1]INESC-ID, Knowledge Discovery and Bioinformatics Group, R. Alves Redol, 9, 1000-029 Lisbon, [2]Instituto Superior Técnico, Technical University of Lisbon and [3]IBB-Institute for Biotechnology and BioEngineering, Centre for Biological and Chemical Engineering, Biological Sciences Research Group, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

Received August 24, 2010; Accepted September 30, 2010

ABSTRACT

The YEAst Search for Transcriptional Regulators And Consensus Tracking (YEASTRACT) information system (http://www.yeastract.com) was developed to support the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. Last updated in June 2010, this database contains over 48 200 regulatory associations between transcription factors (TFs) and target genes, including 298 specific DNA-binding sites for 110 characterized TFs. All regulatory associations stored in the database were revisited and detailed information on the experimental evidences that sustain those associations was added and classified as direct or indirect evidences. The inclusion of this new data, gathered in response to the requests of YEASTRACT users, allows the user to restrict its queries to subsets of the data based on the existence or not of experimental evidences for the direct action of the TFs in the promoter region of their target genes. Another new feature of this release is the availability of all data through a machine readable web-service interface. Users are no longer restricted to the set of available queries made available through the existing web interface, and can use the web service interface to query, retrieve and exploit the YEASTRACT data using their own implementation of additional functionalities. The YEASTRACT information system is further complemented with several computational tools that facilitate the use of the curated data when answering a number of important biological questions. Since its first release in 2006, YEASTRACT has been extensively used by hundreds of researchers from all over the world. We expect that by making the new data and services available, the system will continue to be instrumental for yeast biologists and systems biology researchers.

INTRODUCTION

The model eukaryote *Saccharomyces cerevisiae*, whose genome sequence is available since 1996 (1), plays an essential role in our efforts to understand complex biological networks that control cellular processes. Since the release of the complete yeast genome sequence, a number of computational methods and tools have become available to support research related with this organism. Since its release in 2006, the YEASTRACT (YEAst Search for Transcriptional Regulators And Consensus Tracking; http://www.yeastract.com) (2) database makes publicly available up-to-date information on documented regulatory associations between transcription factors (TFs) and target genes, as well as between TFs and DNA-binding sites, in *S. cerevisiae*. It also periodically integrates the existing data on gene regulation with

*To whom correspondence should be addressed. Tel: +351 213100384; Fax: +351 212145843; Email: atf@inesc-id.pt

data from SGD (3) and the Gene Ontology (GO) Consortium (4). Additionally, a set of bioinformatics tools is also provided to facilitate the full exploitation of the curated data. Although other databases have also made available information about regulatory mechanisms in yeast and other organisms [e.g. MYBS (5) and TRANSFAC (6)] or computational tools for the analysis of promoter regions [e.g. RSAT (7)], YEASTRACT is the information system that most seamlessly integrates extensive regulation data and computational tools for the analysis of this information.

It is worth noticing that biological information systems are getting more specialized each year. This means that the era of replicating information through several biological databases is coming to an end. The trend is creating information systems that obtain the specific information in real time, as they need it, directly from the database which is specialized in that domain and always up to date. This implies that, eventually, all important biological information systems will provide a machine readable interface to access the database. The availability of the implemented resources in YEASTRACT is a clear step forward towards the integration of the biological information about cellular processes.

In this new release, we provide a set of web service resources accessible through a RESTful JSON (JavaScript Object Notation) (8) API (Application Programming Interface) (9). These resources help the user close the gap between the provided functionalities of YEASTRACT web interface and the particular user-tailored queries that are necessary for more advanced uses of YEASTRACT. Each user can access the curated data by developing code on the client side, to query, retrieve and exploit the data in accordance with specific needs.

Another important feature of this release is the improvement of the functionality that enables the user to group a given list of genes by their regulating TFs. At present, the results obtained can be manipulated using a new interactive interface that is useful to help in the understanding of the regulations displayed. The interactive interface is based on a client-side program, written in ActionScript (10), developed with the aid of the Prefuse Flare toolkit (11), and uses the new web service interface to retrieve all the necessary data.

Finally, the information stored in the database has been continuously updated, exhibiting today 80% more regulatory associations than its former release in 2008. Furthermore, the information on regulatory associations has been revised and completed to include detailed knowledge on the experimental evidences that support it.

DATA UPGRADE

Throughout the period of 3 years since the last YEASTRACT release (12), the information in the database has been regularly updated. Since 2008, 20 433 regulatory associations were described in approximately 200 new papers in peer-reviewed international journals and, subsequently, added to the database. The database presently contains more than 48 200 regulatory associations between genes and TFs, based on more than 1200 bibliographic references.

Furthermore, an upgrade in the quality and classification of the stored data was achieved. The 1235 articles that underlie the data gathered so far were revisited by the YEASTRACT curators and information on the precise experimental basis of the associations between TFs and target genes and between TFs and DNA-binding sites were searched for and added to the database. Experimental evidences were, then, classified as direct or indirect. 'Direct Evidence' was considered to be provided through experiments such as Chromatine ImmunoPrecipitation (ChIP), ChIP-on-chip and Electrophoretic Mobility Shift Assay (EMSA), that prove the direct binding of the TF to the target gene's promoter region, or such as the analysis of the effect on target-gene expression of the site-directed mutation of the TF binding site in its promoter region, which strongly suggests that the TF interacts with that specific target promoter. The classification 'Indirect Evidence' was attributed to experiments such as the comparative analysis of gene expression changes occurring in response to the deletion, mutation or over-expression of a given TF. Based on this classification, YEASTRACT contains 29 051 regulatory associations based on direct evidences and 19182 on indirect evidences.

This classification based on different types of experimental evidences allowed the inclusion of additional constraints in the queries *GroupbyTF*, *SearchforTFs* and *SearchforGenes*. For example, it is now possible to group a list of genes (e.g. the genes found to be co-regulated in a given microarray analysis) based on the TFs known to directly interact with their promoter region, excluding those TFs for which there is only indirect evidence of association with the genes in the list. This can prove extremely useful to narrow down a given search or grouping to the TFs that actually count.

WEB SERVICES INTERFACE

YEASTRACT provides curated biological data which is periodically maintained and made available through a set of queries useful to the yeast community. However, one can always imagine many other more complex types of queries or views of the biological data, which can be used to extract useful information. Since the YEASTRACT release in 2006, we have received many requests for the implementation of specific queries in the web interface or for their results to be sent out in flat files. Some of these requests were identified as being very useful to the yeast research community and to the systems biology research community, and therefore implemented and made available. However, many of these requests could not be satisfied, either due to the lack of resources to implement them or to the fact that they would be useful only to a very particular group of researchers.

Available resources

In order to answer the growing demand for tailored requests, YEASTRACT now provides WEBSERVICES. WEBSERVICES is a set of resources (http://www.yeastract.com/services/) for accessing the curated biological data through RESTful web services. It provides the user with a powerful technology to access the YEASTRACT data, bypassing the web interface. Users can therefore develop their own views of the data, by retrieving and exploiting them according to their needs. WEBSERVICES provides the output for the implemented resources through a RESTful JSON API (8,9). The set of resources made available through this API for this release of the YEASTRACT information system is described in Table 1.

The user can find complete documentation on the listed resources as well as a tutorial with a client source code example, in the website http://www.yeastract.com/services/. However, for illustrative purposes, we include in this article a succinct description of one of the resources, namely *GroupbyTF*.

The *GroupbyTF* resource enables the user to retrieve a list of genes grouped according to their documented or potential regulators. The regulators are specified by sending arguments in the call of the web services. As main arguments, the user provides a list of co-regulated genes and a list of the corresponding TF regulators. One can also choose if the regulations are to be restricted to the potential or to the documented regulations. Within the documented regulations, the user can refine the query by

Table 1. List of implemented resources made available through the YEASTRACT web services interface

Resources	Description
Proteins	Retrieves the information relative to a given protein or list of proteins
Loci	Retrieves the information relative to a given ORF/gene or a list of ORF/genes
Regulations	Retrieves regulations between a set of TFs and a set of target genes
GroupbyTF	Retrieves a list of genes grouped according to their documented or potential regulators
References	Retrieves the set of bibliographic references supporting: - the regulations between a set of TFs and a set of target genes - the TF binding sites associated to a given TF

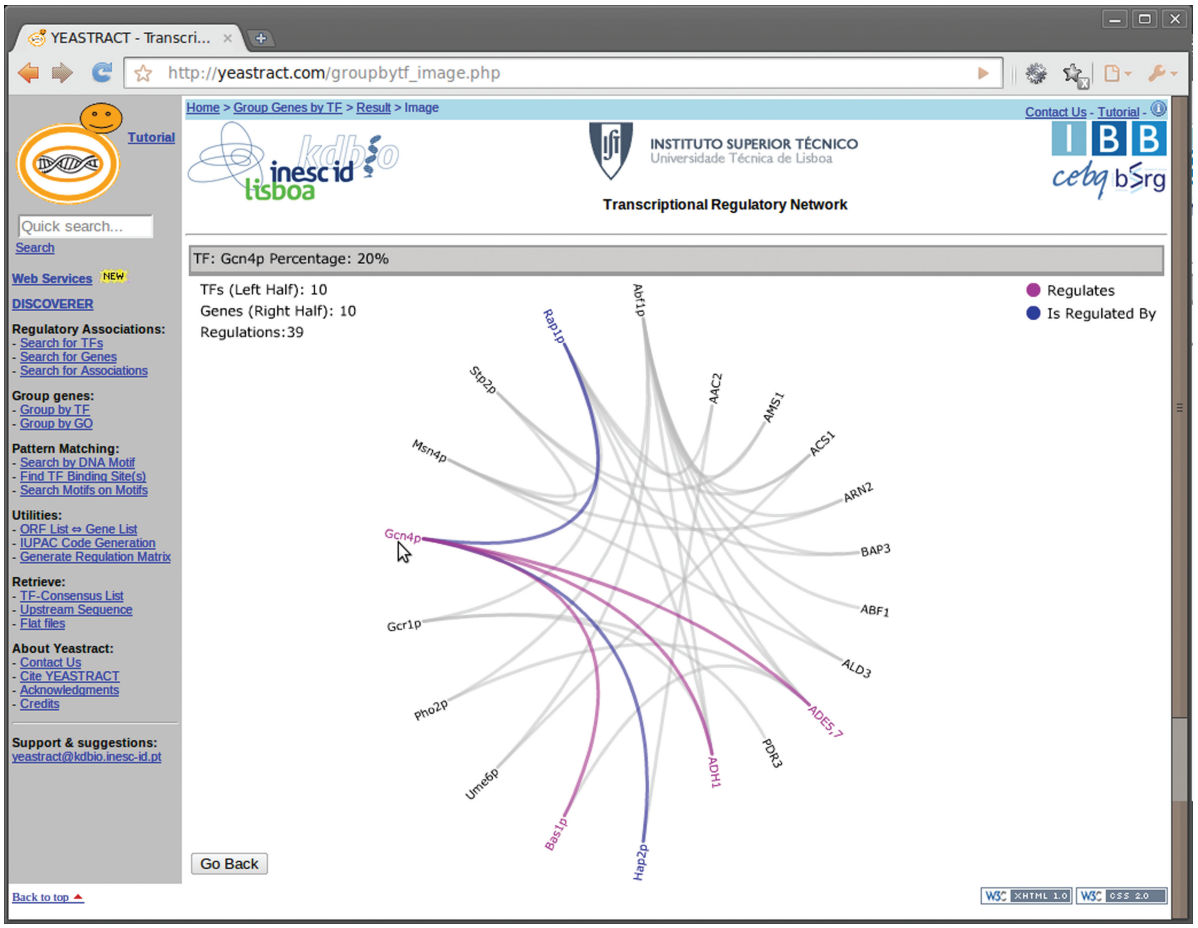

Figure 1. Sample page showing an example output presenting the developed interactive image of the transcriptional regulations between a set of TFs and a set of target genes. This image is generated by a client side application that uses a web service call to the resource *GroupbyTF*.

selecting the type of evidences supporting those documented regulations. The user can also choose if the percentage value representing the proportion of genes regulated by each TF is calculated relative to the total number of genes in the given list or to the number of genes, in the whole yeast genome, documented as being regulated by that TF.

As an example of the use of the *GroupbyTF* resource, the existing *Group genes: Group by TF* query in the YEASTRACT web interface was extended. This query provides, as a result, a table containing a list of genes grouped by their regulating TFs, sorted by showing first the TFs regulating the larger number of genes. In the right side of this table, the user can select the TFs that should be considered in the image generation functionality. In previous releases of the YEASTRACT information system only a static image was available, since extensive manipulation on the server side is cumbersome. This former image format made it virtually impossible to visualize in a single frame regulatory networks comprising more than 4/5 TFs and 20/30 target genes. In this new release, an interactive image can be generated, containing the result of this query. The application uses the *GroupbyTF* web services resource, presenting the result as a circle of regulated genes and regulating TFs (Figure 1). The user can select any of the nodes in the circle to highlight the edges of the corresponding regulations. A blue edge indicates that the selected node is being regulated by the nodes of the incoming edges. A purple edge indicates that the selected node is regulating the nodes of the outgoing edges. Additionally, upon the selection of a TF the corresponding percentage of regulated ORF/genes is indicated, according to the percentage shown in the table that resulted from the query. This new visualization format has also been optimized to make possible the graphical observation of complex networks comprising up to one thousand elements.

FUTURE DIRECTIONS

The YEASTRACT system has been instrumental for yeast biologists and systems biology researchers in the analysis and modeling of transcriptional regulatory networks in *Saccharomyces cerevisiae* (13–15). The equivalent curated biological information currently available in YEASTRACT for *S. cerevisiae* cannot be found, in the scientific literature, for any other yeast species. However, a constantly growing number of complete eukaryotic genome sequences allow computational biologists to explore comparative genomic approaches to predict *cis*-acting regulatory elements and to reconstruct transcriptional regulatory networks or even small regulatory modules.

In this sense, the new strategic objective of the YEASTRACT project is to include in the system the equivalent available information on other yeasts of biotechnological or clinical interest with well annotated genome sequences. This will allow the analysis of transcription regulatory associations also in these yeasts and the reconstruction of regulatory networks based on the existing documented regulatory networks in *S. cerevisiae* and comparative genomics.

ACKNOWLEDGEMENTS

The information about yeast genes other than documented regulations, potential regulations and the transcription factor binding sites contained in YEASTRACT was gathered from SGD and the GO Consortium. The authors are grateful to colleagues and friends from the yeast community for their encouragement and suggestions.

FUNDING

Fundação para a Ciência e a Tecnologia and the PTDC program (projects PTDC/BIO/72063/2006, PTDC/EIA/71587/2006 and PTDC/EIA/67722/2006). Funding for open access charge: PTDC program, projects PTDC/EIA/71587/2006 and PTDC/EIA/67722/2006.

Conflict of interest statement. None declared.

REFERENCES

1. Goffeau,A., Barrell,B., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J., Jacq,C., Johnston,M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 563–567.
2. Teixeira,M.C., Monteiro,P., Jain,P., Tenreiro,S., Fernandes,A.R., Mira,N.P., Alenquer,M., Freitas,A.T., Oliveira,A.L. *et al.* (2006) The YEASTRACT database: a tool for the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **34**, D446–D451.
3. Hong,E.L., Balakrishnan,R., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R., Fisk,D.G., Hirschman,J.E., Livstone,M.S. *et al.* (2007) Saccharomyces Genome Database ftp://ftp.yeastgenome.org/yeast/.
4. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
5. Tsai,H.K., Chou,M.Y., Shih,C.H., Huang,G.T.W., Chang,T.H. and Li,W.H. (2007) MYBS: a comprehensive web server for mining transcription factor binding sites in yeast. *Nucleic Acids Res.*, **35**, W221–W226.
6. Wingender,E., Chen,X., Fricke,E., Geffers,R., Hehl,R., Liebich,I., Krull,M., Matys,V., Michael,H. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, **29**, 281–283.
7. van Helden,J. (2003) Regulatory sequence analysis tools. *Nucleic Acids Res.*, **31**, 3593–3596.
8. Richardson,L. and Ruby,S. (2007) RESTful Web Services. O'Reilly Media.
9. Crockford,D. (2006) The application/json Media Type for JavaScript Object Notation (JSON), http://tools.ietf.org/html/rfc4627.
10. Moock,C. (2007) Essential ActionScript 3.0. O'Reilly Media/Adobe Dev Library.
11. Heer,J., Card,S.K. and Landay,J.A. (2005) Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, pp. 421–430.
12. Monteiro,P., Mendes,N., Teixeira,M.C., d'Orey,S., Tenreiro,S., Mira,N.P., Pais,H., Francisco,A., Carvalho,A. *et al.* (2008) YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **36**, D132–D136.
13. Orlando,A.D., Lin,C.Y., Bernard,A., Wang,J.Y., Socolar,J.E.S., Iversen,E.S., Hartemink,A.J. and Haase,S.B. (2008) Global

control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, **453**, 944–947.
14. Gertz,J., Siggia,E.D. and Cohen,B.A. (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, **457**, 215–218.
15. Teixeira,M.C., Dias,P.J., Monteiro,P.T., Oliveira,A.L., Freitas,A.T. and Sá-Correia,I. (2010) Refining current knowledge on the yeast FLR1 regulatory network by combined experimental and computational approaches. *Molecular BioSystems*, DOI:10.1039/C004881J.